REMARKS

Claims 1-3, 6-12, 14-19, and 21-23 are all the claims presently pending in the application. Claims 4, 5, 13, and 20 are canceled and new claims 21- 23 are added.

It is noted that the claim amendments are made only for more particularly pointing out the invention, and not for distinguishing the invention over the prior art. Further, Applicant specifically states that no amendment to any claim herein should be construed as a disclaimer of any interest in or right to an equivalent of any element or feature of the amended claim.

The Examiner objects to the drawing for allegedly failing to show the feature described in the independent clams that data is streamed from two of three matrices ... to a first matrix. In response, Applicants direct the Examiner's attention to Figure 5, wherein is clearly shown that matrix data 502, 503 are streamed into L1 from L2 and that matrix data 504 from matrix 501 resides in L1. Therefore, Applicants respectfully request that the Examiner reconsider and withdraw this objection.

The Examiner also objects to the previous amendment for allegedly introducing new matter by reason that the Examiner considers that there is no clear support in the original disclosure for streaming data from two of three matrices. In response, Applicants direct the Examiner's attention to the discussion relative to Figure 5 and particularly to the final full paragraph on page 9 of the specification: "*Thus, in the above paragraph, the size of N2 is chosen so that the cache resident piece and the two streaming pieces can "fit" into the TLB. By doing so, TLB thrashing can be avoided.*" Accordingly, Applicants submit that the previous claim revisions are indeed clearly supported in the original disclosure, and the Examiner is respectfully requested to reconsider and withdraw this objection.

The Examiner objects to claims 8, 12, 17, 18, and 20 for failing to define "BLAS." Although Applicants believe that one having ordinary skill in the art would understand this terminology, Applicants have amended these claims to expedite prosecution. Therefore, Applicants respectfully request that the Examiner reconsider and withdraw this rejection.

Claims 1-20 stand rejected under 35 U.S.C. § 112, second paragraph, as allegedly being indefinite. Claims 1-20 stand rejected under 35 U.S.C. § 101 as allegedly directed to non-statutory subject matter. Claims 1-6, 9, 14, 15, and 18 stand rejected under 35 U.S.C. § 102(b) as anticipated by U.S. Patent No. 5,099,447 to Myszewski. Claims 7, 8, 11, 12, 16, and 17 stand rejected under 35 U.S.C. § 103(a) as unpatentable over Myszewski,

further in view of US Patent 6,675,106 to Keenan et al. Claim 19 stands rejected under 35 U.S.C. § 103(a) as unpatentable over Myszewski, further in view of non-patent literature by Philip Alpatov, et al., and claim 20 stands rejected under 35 U.S.C. § 103(a) as unpatentable over Myszewski, further in view of Alpatov and Keenan.

These rejections are respectfully traversed in the following discussion.

I. THE CLAIMED INVENTION

The claimed invention, as exemplarily defined in independent claim 1, is directed to a method of improving at least one of speed and efficiency when executing a linear algebra subroutine on a computer having a memory hierarchical structure including at least one cache. For a level 3 matrix multiplication processing, it is determined which matrix will have data for a submatrix block residing in a lower level cache of the computer and which two matrices will have data for submatrix blocks residing in at least one higher level cache or a memory. The data from the selected two matrices is streamed for the executing of the level 3 matrix multiplication processing.

The present inventors have recognized that conventional linear algebra processing based on LAPACK subroutines, for example, are not optimal.

The claimed invention, on the other hand, along with various other techniques described in the co-pending applications, provides techniques that improve processing efficiency. More specifically, the present invention provides a memory management method allowing for a streaming of data through a cache, using another operand as having the "matrix role" and being resident in the cache.

II. THE REJECTION UNDER 35 USC §112, SECOND PARAGRAPH

The Examiner rejects to claims 1, 9, 14, and 19 for having only one step of streaming, which the Examiner considers not to be sufficient for executing a linear algebra subroutine as described in the preamble.

Although Applicants do not necessarily agree with the Examiner's position, to expedite prosecution, Applicants have revised these claims to more clearly reflect that the invention is directed to improving speed and efficiency in the processing, rather than the

processing itself.

The Examiner also rejects claims 18 for an antecedent problem with the L1 cache. Applicants believe that the above claim amendments appropriately address this concern.

Therefore, Applicants respectfully request that the Examiner reconsider and withdraw this rejection.

III. THE 35 USC §101 REJECTION

Claims 1-20 continue to stand rejected under 35 U.S.C. §101. The Examiner Do characterizes that claims 1-20 "... merely disclose steps of streaming data from cache without [regard] to any particular practical application or tangible result." Examiner Do also considers that claims 14-18 are directed to "signal medium" and, therefore, non statutory.

Applicants again respectfully disagree, since the standard applied to the evaluation of software-related patents is whether the claimed invention as a whole provides a useful, concrete and tangible result. The present invention clearly satisfies this standard since, as a whole, it improves speed/efficiency for processing level 3 matrix multiplication. This result is a tangible result for a computer executing this processing and clearly provides a practical application. The Examiner's confusion seems to be caused by the invention as being related to the processing of matrix subroutines without identifying a specific application for which the matrix subroutines are being applied.

However, the Examiner's rationale overlooks that the practical result of the present invention involves its effect on the machine doing the processing, which result is inherently practical. The present invention is not attempting to claim a mathematical algorithm in the abstract. Nor does the present invention in any way preempt a mathematical algorithm, since the matrix subroutines, is these are considered as "mathematical algorithms", can still be executed without the method of the present invention.

First, relative to claims 14-18, Applicants submit that these claims are clearly addressed to "[a] machine-readable storage medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus ...." These claims are clearly "Beauregard claims", after *In re Beauregard*, 53 F.3d 1583 (Fed. Cir.,1995) and clearly statutory subject matter. The Examiner is respectfully requested to

identify specific case law in support of any continued rejection.

Relative to claims 9-13, these claims are directed to an apparatus and are, therefore, clearly directed to a machine, another of the four enumerated categories specifically recited in 35 USC §101 as statutory subject matter.

Relative to the method claims, the present invention as a whole is directed to memory management of data as used for the specific application of processing linear algebra subroutines. As such, the invention clearly relates to improvement of speed and efficiency of a computer executing these subroutines. To the extent that such improvement in computer processing requires justification under the "useful, concrete and tangible result" test, Applicants submit that the improvement of speed and efficiency and the mechanism of data management inherently satisfy this test.

As discussed during the telephone interview on March 14, 2007, the present invention does indeed have the prerequisite practical application and tangible result (wherein "tangible" means "real-world") because its method improves DGEMM performance in two ways: faster DGEMM kernel processing via streaming; and, providing only 4 ways to block for the entire DGEMM operation via the use of faster DGEMM kernels. Since these two benefits provide improved efficiency and speed in DGEMM processing on a computer, there clearly is a beneficial real world result.

Thus, from another perspective, the present invention can be viewed as a memory management technique in a computer that improves efficiency for processing dense linear algebra subroutines.

Should the Examiner wish to maintain this rejection, Applicants request that specific case law be cited so that Applicants can make an appropriate comparison of the present invention with the facts of the facts behind that case law holding.

In view of the foregoing, the Examiner is respectfully requested to reconsider and withdraw this rejection.

IV. THE PRIOR ART REJECTIONS

The Examiner alleges that Myszewski teaches the claimed invention defined by claims 1-6, 9, 14, 15 and 18, when modified by Keenan, renders obvious claims 7, 8, 11, 12, 16, and 17, and, when modified by Alpatov, renders obvious claim 19. The Examiner

also alleges that further modification of Mszewski/ Alpatov, by Keenan renders obvious claim 20.

Applicants respectfully disagree and submit that there are elements of the claimed invention which are neither taught nor suggested by Myszewski.

The present invention provides a method and structure for producing high performance linear algebra routines using streaming. An exemplary feature of this invention is that it allows an L1 cache resident matrix to get essentially infinite reuse. Hence, the term "streaming" is used, although this choice of terminology may be unfortunate, since "streaming" has other connotations in this field. Also, by using streaming in accordance with the present invention, one can reduce the number of efficient ways to block data for matrix multiplication.

In contrast, Myszewski is using standard prior-art cache blocking techniques for DGEMM alone and mostly for the Alliant FX/2800 computer. The present invention is concerned with a generalization of level 1 cache blocking which is applicable to almost all level 3 dense linear algebra (L3 DLA) algorithms. A key aspect of the present invention is that the processing initially determines which of the three matrices is smallest in size and should be resident in L1 cache and which two remaining matrices should be streamed from higher levels.

Myszewski has no concept corresponding to this initial selection of which matrix will be cache resident and which two matrices will be streaming from higher levels. In fact, Myszewski always uses the A matrix of the three to be resident in the L1 cache and hence, clearly teaches against the concept of the present invention.

Furthermore, the present invention has only one of three matrices A, B, C in cache while a vector block and a scalar block of the remaining two matrices are being streamed from the next cache level. Since the next cache level is much larger, the current block in cache get a much greater reuse factor. In contrast, Myszewski is using standard cache blocking and therefore does not have this feature.

The present invention works for cache level i+1 and i where i runs over L-1 cache level and memory for an L level memory heirarchy. In contrast, Myszewski only addresses a single cache and memory.

The present invention is choosing a matrix argument that is small relative to the other two matrix arguments. Streaming requires that the other two matrix arguments be

large (conservation of matrix data).

Applicants submit that this discussion demonstrates that there are clear differences between Myszewski and the present invention, particularly in the fact that the present invention generalizes the streaming concept to select which of the matrices should be streamed. The Examiner relies upon the secondary references for reasons unrelated to this basic deficiency of Myszewski, so the secondary references fail to overcome this basic deficiency.

Hence, turning to the clear language of the claims, in Myszewski there is no teaching or suggestion of: "…determining, for a level 3 matrix multiplication processing, which matrix will have data for a submatrix block residing in a lower level cache of said computer and which two matrices will have data for submatrix blocks residing in at least one higher level cache or a memory; and streaming data from said selected two matrices in executing said level 3 matrix multiplication processing ….", as required by independent claim 1. The remaining independent claims have similar language. Therefore, all claims are clearly patentable over Myszewski.

Therefore, Applicants submit that there are elements of the claimed invention that are not taught or suggest by Myszewski. Therefore, the Examiner is respectfully requested to withdraw this rejection.

V. FORMAL MATTERS AND CONCLUSION

In view of the foregoing, Applicants submit that claims 1-3, 6-12, 14-19, and 21-23, all the claims presently pending in the application, are patentably distinct over the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.

Should the Examiner find the application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary in a telephonic or personal interview.

The Commissioner is hereby authorized to charge any deficiency in fees or to credit any overpayment in fees to Assignee's Deposit Account No. 50-0510.

Respectfully Submitted,

Date: July 16, 2007

Frederick E. Cooperrider
Registration No. 36,769

McGinn Intellectual Property Law Group, PLLC
8321 Old Courthouse Road, Suite 200
Vienna, VA 22182-3817
(703) 761-4100
Customer No. 21254